Dean&Francis
ISSN 2959-6157

Deep Reinforcement Learning in Video Games

Junren Lu

Shanghai Film Academy, Shanghai University, Shanghai, 200072, China
*Corresponding author: reien@shu.edu.cn

Abstract:

The high interactivity and instant feedback characteristics of games are highly compatible with the trial-and-error learning mechanism of Deep Reinforcement Learning (DRL). In recent years, DRL has achieved a series of landmark breakthroughs in Game AI, from Atari games to superhuman levels in complex multi-agent game environments. With the integration of Large Language Models (LLMs) with DRL, especially Multi-Agent Reinforcement Learning (MARL), DRL applications scenarios are no longer limited to performance score, but have gradually evolved into more complex situations requiring social reasoning and human-machine collaboration. This paper reviews the application of DRL methods in game AI with some famous games, which can be classified into three categories of games according to the type of games, such as classical Arcade games, First-Person 3D games and Multiplayer Competitive games. This paper conducts a detailed review of the evolution of DRL approaches. In addition, there are some challenges existing in this field, like AI agents' generalization and human-machine collaboration capabilities. This paper also discusses current multi-agent games and some emerging methods that combine LLM with MARL, while outlining future research directions. This review aims to provide researchers with a clear technical overview.

Keywords: Deep Reinforcement Learning; Game AI; Reinforcement Learning; Multi-Agent Reinforcement Learning; Large Language Models.

1. Introduction

From early board games to modern complex 3D games, video games have offered numerous environments for AI research. Due to video games' clear rules, quantifiable objectives and variety of types, video games have always been an essential application domain for machine learning research [1]. Game environments evaluate an algorithm's ability of decision-making, planning, learning and interacting with multiple agents. In recent years, one of the most powerful techniques in the area of game AI is Deep

Reinforcement Learning (DRL) due to several breakthroughs in Deep Learning (DL).

By combining the representation learning capabilities of DL and the decision optimization capabilities of Reinforcement Learning (RL), Deep Reinforcement Learning allows AI agents to learn and act directly from high-dimensional inputs, e.g., pixels. In 2013, DeepMind proposed Deep Q-Network (DQN) that can learn directly from pixel inputs. DQN achieved human-level performance on several Atari 2600 games [2]. This result opened a new stage for DRL in game AI research, which shows the possibility of end-to-end learning in complex environments.

Afterwards, DRL also gained several breakthroughs in game AI applications. In 2016, AlphaGo beat world Go champion Lee Sedol using a new method that integrated Monte Carlo Tree Search (MCTS) into Deep Neural Networks [3]. In 2017, AlphaGo Zero beat all the previous versions without any human data by self-play [4]. In 2019, OpenAI Five beat professional players in Dota 2 using DRL [5] and AlphaStar reached Grandmaster level in StarCraft II [6].

In recent years, with the achievement of Large Language Models (LLM), DRL applications in video games have expanded from pure performance optimization to scenarios requiring social reasoning and social environments. Researchers have begun to try to combine language reasoning ability with DRL, such as combining Multi-Agent Reinforcement Learning (MARL) with LLM, impressive strong performance in social deduction games like Among Us [7]. This direction demonstrates DRL's potential in more complex, human-like social settings.

This survey intends to systematically review the application of DRL in video games. It will introduce the evolution and implementation of DRL methods in different game types, and discuss some existing challenges and future trends.

2. Fundamental Theory

Deep Learning is mainly used for data representation. Common networks include: Convolutional Neural Networks (CNN), dominating in computer vision; Recurrent Neural Networks (RNN), a particular type of Deep Neural Network suited for Natural Language Processing; Long Short-Term Memory Networks (LSTM), a specialized form of RNN capable of dealing with sequence-dependent tasks.

Reinforcement Learning is a method where an agent learns optimal policies through trial and error by interacting with an environment and receiving a reward signal. The Markov Decision Process (MDP) provides the mathematical framework for RL. MDP consists of five core elements: states (S), actions (A), transition probabilities (P), rewards (R), and a discount factor (γ). Together, these components define the structure of an MDP as a tuple $\langle S, A, P, R, \gamma \rangle$.

When applied to video games, the game itself serves as the environment. The agent, acting as a player, takes a finite set of actions at each step, with reward signals determined by the game score. In video games, reward signals may occur frequently or rarely. For instance, Open-World Games often lack explicit reward models. Therefore, when applying RL to games with sparse rewards, the reward R(s) for states must propagate back to the preceding actions that led to the reward. Consequently, delayed sparse reward scenarios pose significant challenges for RL algorithms [8].

2.1 Deep Reinforcement Learning

The integration of DL and RL has given rise to DRL. DRL is primarily categorized into three types of algorithms: value-based, policy gradient and model-based DRL methods.

Value-based DRL: The most classic value-based DRL method is Deep Q-network (DQN). It directly learns policies by taking raw pixels as input and producing value functions to estimate future rewards. Two approaches address the training instability, i.e., Experience Replay and Target Networks.

Policy Gradient DRL: Actor-Critic is a combination of value-based method and policy gradient method. It uses a value-based criterion function to compute policy gradients for estimating expected future returns. Asynchronous DRL further develops asynchronous gradient descent for the policy optimization [9]. Asynchronous Advantage Actor-Critic (A3C) stands out by training multiple agents in a variety of different environments, with stable training performance.

2.2 Multi-Agent Reinforcement Learning

In many real-world environments and game environments, there are multiple agents existing and needed to coexist with each other. Multi-agent environments introduce two fundamental challenges, credit allocation and environmental non-stationarity. Credit Allocation is the problem of determining each agent's contribution to the collective outcome. Environmental non-stationarity is caused by the simultaneous learning and adaptation of all agents, which makes the environment non-stationary from any individual agent's perspective. Consequently, MARL extends the RL research to solve the challenge of multiple agents learning together and interacting in a shared environment.

ISSN 2959-6157

According to the communication pattern, MARL can be mainly classified into three major types: Fully Decentralized methods without any central coordination, Fully Centralized methods with unified control, and hybrid frameworks like Centralized Training with Decentralized Execution (CTDE [10]).

Fully decentralized: All agents learn their policies in an independent way, and all rely on their own local observations and rewards. This approach is essentially single-agent RL, where other agents are treated as part of the environment. However, the dynamic changes in their policies lead to environmental non-stationarity, severely undermining the Markov stationarity assumption required for convergence in traditional RL algorithms (such as Q-learning) [11]. Although this approach is simple and requires no inter-agent communication, its poor stability makes it difficult to converge to optimal cooperative policies. A typical algorithm is Independent Q-learning (IQL) [12].

Fully Centralized: Aggregates all agents' local observations into a global state and treats the joint action space as a single action for training by a central controller, thereby transforming the MARL problem into a massive single-agent RL problem. Theoretically, this approach can find optimal solutions, but its decision space grows exponentially with the number of agents. Therefore, it is impractical for many real-world applications requiring rapid distributed responses.

CTDE: CTDE is regarded as the most successful MARL approach in recent years, ingeniously combining the strengths of the other two methods. During training, the system leverages global information (such as the global state, actions or policies of other agents) to learn a centralized Critic or perform efficient gradient updates, thereby addressing credit allocation and environmental non-stationarity issues. At execution time, each agent executes its own policy independently, acting solely based on its local observations, enabling fully decentralized deployment [4].

3. The evolution of DRL in Video Games

Video games can be categorized into various types based on their gameplay: Classic Arcade Games (e.g., Atari 2600 games), First-Person games (e.g., ViZDoom), Real-Time Strategy (RTS) games (e.g., StarCraft II), turn-based strategy and board games (e.g., Go, chess, card games), and Multiplayer Competitive games (e.g., DOTA2, text-based social deduction games). This part reviews the primary applications and evolution of DRL across these game types.

3.1 Classical Arcade Games

Atari 2600 games served as the standard testing platform for early DRL development, offering relatively simple visual inputs and discrete action spaces ideal for algorithm evaluation. DQN first performed as good as human players across multiple Atari games in 2013, pioneering the end-to-end learning paradigm from pixels to actions [2].

Subsequently, algorithmic refinements further enhanced performance on Atari games. Double Q-learning (Double DQN) addressed Q-value overestimation by decoupling the selection of the action from the evaluation of that action [13]; Dueling DQN accelerated learning through separate branches for state values and advantage functions, while Prioritized Replay improved sampling efficiency [14]. Ultimately, Rainbow DQN has emerged as a significant advancement by effectively integrating a variety of those methods (Multi-step Learning, Double DQN, Dueling Networks, Noisy Net, Distributed RL, etc.) [15], achieving state-of-the-art performance on Atari benchmarks and establishing itself as the current baseline for DRL algorithms. These developments demonstrate clear trends of DRL evolution, where the focus has shifted from a single optimization objective to multi-dimensional technology integration, and evolved from stability improvements to significant performance breakthroughs.

3.2 First-Person 3D Games

For first-person 3D games (e.g., VizDoom), core challenges include visual complexity and partial observability. Early research revealed that value-based DQN exhibits limited performance in such environments, leading to greater adoption of Actor-Critic methods. For instance, in the 2016 VizDoom competition, the "Clyde" team employed an A3C architecture to achieve competitive results even in complex 3D multi-agent scenarios [16]. That same year, DeepMind's UNREAL algorithm incorporated unsupervised auxiliary tasks like pixel control into A3C, accelerating learning by 10x on 3D tasks like Labyrinth while achieving 880% human baseline performance on Atari tasks [17].

3.3 Multiplayer Competitive Games

StarCraft II is a popular and classic platform for MARL research. This type of RTS game demands simultaneous consideration of macro-level strategy and micro-level operations, exhibiting extremely high dimensional complexity. In 2019, DeepMind's AlphaStar reached Grandmaster level in StarCraft II's global mode by using a Transformer model to apply attention relationships between game units in a multiplayer environment [6]. AlphaStar used multiple DRL techniques—such as deep neural networks, policy

gradients, and adversarial self-play—along with tricks like experience replay and supervised learning pre-training to show that DRL can solve complex real-time multi-agent coordination tasks.

Hanabi is a cooperative card game where players see only each other's hands and coordinate through limited hints. Hanabi Challenge was released by Bard et al. (2019) who claimed that this game places "inference about others' beliefs and intentions" (theory of mind) in prominence [18]. Traditional DRL methods (e.g., Rainbow) perform well in pure self-play environments but have difficulty in hybrid human-agent settings as well as cooperative games where implicit linguistic meaning is important. Therefore, researchers developed approaches like probabilistic reasoning and theoretical models. For instance, Fuchs et al. proposed a DRL method that incorporates finite-level mental reasoning and intrinsic reward mechanisms to allow agents to strategically share critical information and learn more effective collaboration in Hanabi [19]. These works have shown that in cooperative games, algorithms must be able to develop and use belief models about partners' intentions in order to allow their algorithms to perform well in information-implicit environments.

In recent years, social reasoning games with deception, persuasion, and linguistic communication—such as Among Us and Werewolf—have emerged as a new frontier for MARL. With the rise of LLM, more and more researches have developed textual representation multi-agent environments, to allow LLMs to collaborate and compete as agents. This integration addresses key limitations of both. LLMs excel at language processing which is applied to Knowledge Transfer, as well as complex social reasoning, while MARL provides a robust framework for optimizing sequential decision-making and strategic interactions through trial and error. While some works apply multi-agent RL frameworks to LLM cooperative tasks, a study in 2025 on Hanabi textualizes the game to allow knowledge to be transferred across tasks. Textual encoding is more intuitive and allows for stronger state space generalization ability, which is an advantage when learning universal strategies across different game configurations [20]. Specialized platforms have also been designed for Werewolf-like games. These platforms leverage LLM reasoning capabilities to evaluate model performance in attack, defense, inference, and deception. A Study constructed a textual version of Among Us, training human-like, lying models with RWKV language models via zero-shot learning. These models demonstrated robust performance across varying map sizes and player counts [7]. Overall, integrating language models with DRL enables agents to learn within game involving natural language and social strategies. This represents a novel research direction emerging over the past two years, with future potential to expand into more game scenarios requiring communication and reasoning, as well as collaborative trust-based interactions with humans.

4. Challenges and Future Trends in Games with DRL

Knowledge transfer across new tasks, rather than overfitting to specific environments, remains a key challenge for future DRL systems. Developing reinforcement learning capable of cross-game generalization, such as meta-RL, aims to train agents that can rapidly adapt to new tasks by learning universal game principles through training across different games. For example, the Decision Transformer [21] applies Sequence Modeling to RL, demonstrating the ability to transfer knowledge across multiple tasks.

Current DRL systems are mostly designed to replace or compete against human players, yet Human-AI Collaboration is a vital and unavoidable future topic. A study from MIT found that in the cooperative game Hanabi, while AI agents achieved high scores when competing against each other, most human players rated their behavior as "unnatural" and "unpredictable" when playing alongside AI teammates [22]. Future research must explore collaborative methods that better align with human expectations.

The integration of LLM with DRL represents a new hotspot. LLM provides common-sense reasoning and language comprehension capabilities, while DRL offers decision optimization. This combination excels in social games requiring natural language interaction and holds promise for broader environmental applications [7]. Key challenges include computational efficiency and multi-modal learning, which involves coordinating information from diverse modalities like vision, language, and action in games.

5. Conclusions

This paper reviews DRL in various game environments. DRL is constantly innovating and achieving super-human results while advancing from classic Arcade Games to complex Multiplayer Competitive Games. Different types of games place different demands on algorithms and push their developments, such as First-Person Games, Real-Time Strategy games, and Social Deduction games. However, using specific models for specific situations does not equip AI to navigate more complex real-world environments. DRL faces challenges such as cross-game generalization, human-AI collaboration, and integration with Large Language Models. Future research should focus on approaches like MARL with LLM, Knowledge

ISSN 2959-6157

Transferring to address these challenges. With achievements in methods and computational power, DRL holds promise to play a greater role in broader gaming and real-world environments, enabling more intelligent and human-like performance in game AI.

References

[1] Shaheen A., Badr A., Abohendy A., et al. Reinforcement learning in strategy-based and Atari games: a review of Google DeepMind's innovations. arXiv preprint arXiv:2502.10303: 2025.

[2] Mnih V., Kavukcuoglu K., Silver D., et al. Playing Atari with Deep Reinforcement Learning. arXiv preprint arXiv:1312.5602, 2013.

[3] Silver D., Huang A., Maddison C.J., et al. Mastering the game of Go with deep neural networks and tree search. Nature, 2016, 529(7587): 484–489.

[4] Silver D., Schrittwieser J., Simonyan K., et al. Mastering the game of Go without human knowledge. Nature, 2017, 550(7676): 354-359.

[5] Berner C., Brockman G., Chan B., et al. Dota 2 with large scale deep reinforcement learning. arXiv preprint arXiv:1912.06680, 2019.

[6] Vinyals O., Babuschkin I., Czarnecki W.M., et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature, 2019, 575(7782): 350–354.

[7] Sarkar B., Xia W., Liu C.K., et al. Training language models for social deduction with multi-agent reinforcement learning. arXiv preprint arXiv:2502.06060, 2025.

[8] Shao K., Tang Z., Zhu Y., et al. A survey of deep reinforcement learning in video games. arXiv preprint arXiv:1912.10944, 2019.

[9] Grondman I., Busoniu L., Lopes G.A., Babuska R. A survey of actor-critic reinforcement learning: Standard and natural policy gradients. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2012, 42(6): 1291-1307.

[10] Foerster J.N., Chen R.Y., Al-Shedivat M., et al. Learning with Opponent-Learning Awareness. arXiv preprint arXiv:1709.04326, 2017.

[11] Claus C., Boutilier C. The dynamics of reinforcement learning in cooperative multiagent systems. AAAI/IAAI, 1998(746-752): 2.

[12] Tan M. Multi-agent reinforcement learning: Independent vs. cooperative agents. Proceedings of the Tenth International Conference on Machine Learning, 1993: 330-337.

[13] Van Hasselt H., Guez A., Silver D. Deep reinforcement learning with double Q-learning. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1).

[14] Wang Z., Schaul T., Hessel M., et al. Dueling network architectures for deep reinforcement learning. International Conference on Machine Learning. PMLR, 2016: 1995-2003.

[15] Hessel M., Modayil J., Van Hasselt H., et al. Rainbow: Combining improvements in deep reinforcement learning. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1).

[16] Ratcliffe D.S., Devlin S., Kruschwitz U., et al. Clyde: A Deep Reinforcement Learning DOOM Playing Agent. AAAI Workshops, 2017.

[17] Jaderberg M., Mnih V., Czarnecki W.M., et al. Reinforcement Learning with Unsupervised Auxiliary Tasks. arXiv preprint arXiv:1611.05397, 2016.

[18] Bard N., Foerster J.N., Chandar S., et al. The Hanabi Challenge: A New Frontier for AI Research. Artificial Intelligence, 2020, 280: 103216.

[19] Fuchs A., Walton M., Chadwick T., et al. Theory of Mind for Deep Reinforcement Learning in Hanabi. arXiv preprint arXiv:2101.09328, 2021.

[20] Sudhakar A.V., Nekoei H., Reymond M., et al. A Generalist Hanabi Agent. arXiv preprint arXiv:2503.14555, 2025.

[21] Chen L., Lu K., Rajeswaran A., et al. Decision Transformer: Reinforcement learning via sequence modeling. Advances in Neural Information Processing Systems, 2021, 34: 15084-15097.

[22] Siu H.C., Peña J., Chen E., et al. Evaluation of human-AI teams for learned and rule-based agents in Hanabi. Advances in Neural Information Processing Systems, 2021, 34: 16183-16195.